

TRANSCRIPT – Harmonisation for the EPT demonstration project, ECR module

Slide 1

Hello, I am Helen Collins and I am going to talk to you about the harmonisation for the demo project assessing sex differences in perinatal survival in extremely preterm infants.

Slide 2

In this video you will learn about:

- How we defined the populations for each analysis.
- How harmonised variable definitions were developed.
- How we identified cohort variables for harmonisation.

Slide 3

For this project, all RECAP preterm cohorts were invited to take part. 19 cohorts agreed to take part in the project consisting of 18 regional/national cohorts, and the EPICE multinational cohort.

Slide 4

For this study we focused on extremely preterm infants born at 22-26 weeks gestational age. We found that the recruitment points and data collected for the cohorts meant that there were differences in the baselines we could use for our analysis as some cohorts included stillbirths or infants alive at the onset of labour, while other cohorts only included infants who were alive at admission to the neonatal unit. In order to include as many cohorts as possible, we chose to run several analyses using different baselines.

Slide 5

We developed a matrix to examine which of the cohorts could be used for each substudy based on the cohort starting point for data collection and the perinatal survival variables available.

Slide 6

Here you can see the years of birth of the participants from the different cohorts. Changes in antenatal and neonatal care over time mean it was important to consider the age of the cohorts for this project. We wanted to assess the impact of year of birth in our analysis by stratifying data or using time as an interaction term.

Slide 7

For the Nordic register study, there are also changes in data available over time, as data for Norway are available from 1967 but in Finland only from 1987. There have also been improvements in the data collected over time such as enhanced perinatal data collected in Finland from 2006. It is important to consider how these differences may impact the results of the analysis.

Slide 8

For the harmonisation, the first step was to decide on the variables to include in the analysis and these were based on the research question looking at sex differences in perinatal survival. We started by developing an initial target definition for each variable. In many cases we used simple binary yes/no variables as these allowed us to include the most cohorts in our study, but for other analyses it may be important to have a more detailed or stricter definition. We assessed the variables available in each cohort that could be used to construct the target variables and in some cases there was an iterative process of definition development. We tried to balance including as much detail as possible (or as much detail as needed for the analysis), while including as many cohorts as possible.

Slide 9

For surfactant we chose a target definition of administration of any surfactant to the infant as a simple binary variable with categories no, yes and missing. Some cohorts only collected this level of information while others collected more detailed information about timing of surfactant administration so in this case we chose the simplest definition of any surfactant the infant received at any time point during the neonatal period.

Slide 10

To search for variables on the RECAP platform we used the search function. We selected the schema module of interest and selected a domain to search in this case Surfactant Administration. We then looked at the search results and you can see the variables for each cohort. In this case, we are interested in the variable from ACTION f54 and if we click on it, the variable definition and categories are displayed and the summary statistics available.

Slide 11

For each variable in the demonstration project, we assessed whether there were variables from each cohort that matched the initial harmonised variable definition or could possibly be used to construct the harmonised variable. Once we had a list of all the potential variables available across the cohorts, we discussed any queries with the cohorts, and in some cases made slight changes to our target variable definitions. The categories for the harmonised variables are important and should be carefully assessed as in some cases the target variable couldn't be constructed from the available categories even though the definition initially appeared to match.

Slide 12

Once we had developed our final set of harmonised variables, we created a harmonisation dictionary that included the harmonised variable names and definitions we developed through our iterative harmonisation process. The harmonisation dictionary is in the format required by the RECAP Preterm data platform.

It is important to include as much information as possible in the variable definition so it is clear how the variable is to be harmonised.

We sent the list of possible variables to be used for harmonisation for each cohort and the harmonisation dictionary to the cohort lead and data manager along with some suggestions on how the variables could be harmonised and discussed any queries and issues with cohorts. You will see the process of harmonising variables on the platform using harmonisation scripts in the next video.

Slide 13

We found that collaboration and regular discussion between cohorts and project team is vital for the success of the harmonisation. There are often queries about the variable definitions that the cohort leads are best able to answer. It takes a lot of time to develop variable target definitions because this is an iterative process that is initially based on an ideal definition but is refined based on the available data from the cohorts until a final acceptable definition is reached.

The RECAP Preterm data platform facilitates variable searching as variable level metadata is publicly accessible and searchable. You will hear more about how we developed harmonisation scripts and documented the harmonisation on the data platform in a later video.

Slide 14

References and further information are available here.

Slide 15

Thank you for listening.